DOCUMENT RESUME

ED 156 574 SO 010 967

AUTHOR Munsterman, G. T.; And Others
TITLE Multiple Lists for Juror Selection: A Case Study for San Diego Superior Court.
INSTITUTION American Univ., Washington, D.C. Law Inst.
SPONS AGENCY National Inst. of Law Enforcement and Criminal Justice (Dept. of Justice/LEAA), Washington, D.C.
PUB DATE May 78
CONTRACT J-LEAA-013-76
NOTE 78p.; Tables and figures on pages 3-2, 3-3, 6-3, 6-4, A-6, A-7, A-9, A-11, A-12, B-1 may not reproduce clearly in hard copy due to small type size of original document
AVAILABLE FROM Superintendent of Documents, U.S. Government Printing Office, Washington, D.C. 20402 ($2.30, paper cover)

EDRS PRICE MF-$0.83 HC-$4.67 Plus Postage.
DESCRIPTORS Case Studies; Census Figures; *Courts; *Demography; *Personnel; *Selection; *Statistical Data; Voter Registration
IDENTIFIERS California (San Diego); *Juries

ABSTRACT

Since registered voters have not been found to be an inclusive enough pool of possible jurors in some cities, this report attempts to make specific recommendations to the Superior Court of San Diego County for broadening the juror pool. Aspects of the theory and background of list matching are brought together to support and illustrate the recommendations made. Various possible lists such as telephone lists and driver lists that might be used in conjunction with registered voter lists are mentioned with their individual merits and demerits. For San Diego, the report recommends adding only the driver list. The problem of duplicate recognition is discussed and criteria are given for identifying duplicate records from voter and driver lists, random sampling from lists, and the combining of three or more lists. After combining and editing lists some duplicate and nonresident names often remain. The last two sections of the report cover balancing the level of duplicates remaining on the list with the probability of reflecting nonduplicate names, and the use of qualification questionnaires to remove nonresidents from the list. The report concludes with related readings and five appendices which give demographic information and further methodology specifics. (BC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED156574

U S DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Multiple Lists for Juror Selection:

A Case Study for San Diego Superior Court

by
G. T. Munsterman
C. H. Mount
William R. Pabst

May 1978

SP 010 967

National Institute of Law Enforcement and Criminal Justice
Law Enforcement Assistance Administration
U. S. Department of Justice

National Institute of Law Enforcement
and Criminal Justice

Blair G. Ewing, Acting Director

Law Enforcement Assistance Administration

James M. H. Gregg, Acting Administrator

This project was supported by Contract Number J-LEAA-013-76, awarded to the Criminal Courts Technical Assistance Project, Institute for Advanced Studies in Justice, The American University Law School by the Adjudication Division, Office of Regional Operations, Law Enforcement Assistance Administration, U.S. Department of Justice, under the Omnibus Crime Control and Safe Streets Act of 1968, as amended. Points of view or opinions stated in this document are those of the subcontractor, Bird Engineering-Research Associates, Inc., and do not necessarily represent the official position or policies of the U.S. Department of Justice.

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402
Stock No. 027-000-00665-5

CONTENTS

Section		Page
1.	INTRODUCTION	1-1
2.	RECOMMENDATIONS FOR SAN DIEGO COUNTY	2-1
3.	MULTIPLE LISTS TO BE USED	3-1
3.1	Summary of Recommendation One	3-1
3.2	Discussion	3-2
4.	DUPLICATE RECOGNITION	4-1
4.1	Decision Errors	4-1
4.2	Types of Record Errors	4-4
4.3	Recommended Matching Criteria	4-7
5.	METHODOLOGY FOR THE COMBINING OF LISTS	5-1
5.1	Introduction	5-1
5.2	Direct Combination of Lists	5-1
5.3	Random Selection Without Full List Combination	5-2
5.4	Combination of Three or More Lists	5-4
6.	ERROR RECOGNITION USING QUESTIONNAIRE RESPONSES	6-1
7.	GEOCODING	7-1

Appendix A -- Background Into the Use of Multiple Lists

Appendix B -- Memorandum

Appendix C -- Demographic Characteristics of Voter Registration Lists and Motor Vehicle Drivers

Appendix D -- Determination of List Overlaps and Efficiencies

Appendix E -- Worksheet

Related Reading

ABSTRACT

Lists of registered voters provide the principle source of names for selecting prospective jurors in federal and state courts in the United States. However, voter lists vary from jurisdiction to jurisdiction with respect to the balance of the cross section and with respect to the inclusiveness of the population. To overcome these deficiencies, many courts are supplementing the voter lists with other lists, such as the motor vehicle driver lists, telephone lists, utility lists, and others.

This report examines in some detail the lists available in San Diego, California. It recommends that the Superior Court use a combination of the voter and driver lists as a source of names. It also recommends specific technology for combining these two lists. It finds these two lists to be complementary with respect to both balance and inclusiveness. Other available lists are found to have serious weaknesses. The problem of duplicate recognition is discussed and a matching criterion given. A new method using questionnaire responses to reduce the duplicate level while maintaining a low probability of excluding a good name is given.

Courts now using multiple lists combine the entire list and then select only a few names as prospective jurors. A recently developed technique to achieve equal probability of selection without full list combination is discussed and the methodology illustrated. This method is shown to save a great deal of computer and/or personal time. The problem of geocoding names into proper court jurisdictions is also discussed. Appendices to the report provide a general background to the current use of multiple lists and a discussion of the list characteristics.

While being a technical assistance report addressing the specific situation in San Diego, the methodology applies to any jurisdiction considering the use of multiple lists.

ACKNOWLEDGEMENT

Bird Engineering-Research Associates, Inc., acknowledges with great appreciation the cooperation and assistance of the Office of the Jury Commissioner of The Superior Court for San Diego County and the Jury Committee of that court.

Operating experience was provided by the following and is gratefully acknowledged:

Office of Technical Operations, State of Alaska, Anchorage, Alaska

Office of the Court Administrator and Jury Commissioner, Superior Court, San Mateo County, California

Judicial Department, State of Colorado, Denver

Federal District Court, Denver, Colorado

Fourth Judicial Circuit, Ada County, Boise, Idaho

Office of the Court Administrator, Wyandotte County Courts, Wyandotte, Kansas

Ward County Court, Minot, North Dakota

Professors J. B. Kadane and J. P. Lehoczky of the Department of Statistics, Carnegie-Mellon University, Pittsburgh, Pennsylvania, served as consultants to this project and are responsible for many of the innovations developed in the use of multiple lists.

Special thanks are due the Court of Burleigh County, Bismarck, North Dakota, for their testing of the method for random selection without full list combination.

G. T. Munsterman, Project Director
C. H. Mount, Associate
W. R. Pabst, Jr., Senior Consultant

Publication Note

This is a republication of a technical assistance report originally published by the American University at the request of the Adjudication Division, Office of Regional Operations, Law Enforcement Assistance Administration.

THE USE OF MULTIPLE LISTS FOR JUROR SELECTION

A REPORT TO THE SUPERIOR COURT OF SAN DIEGO

1. INTRODUCTION

The California statute on juror selection was revised in 1975 (Section 204e. California Code of Civil Procedure, Supplement 1976) allowing jurisdictions to supplement the list of registered voters with the list of licensed drivers. The San Diego Superior Court, recognizing the uneven voter registration among the population, felt it necessary to consider use of the wider source list afforded by the statute. The question was whether to do it and how to do it. No guidance as to the methods or procedures was provided by the code. The subject of using multiple lists for juror selection was briefly discussed in A Guide to Jury System Management. The court therefore turned to the Criminal Courts Technical Assistance Project of American University. The Project contracted with Bird Engineering-Research Associates, Inc., to explore the subject and to make recommendations to the Superior Court of San Diego County.

Exploration of the subject matter necessarily went beyond the particular problems of San Diego County. The experience of states and jurisdictions using two or more lists for juror selection revealed the great variety of methods and lists that are in common use. This background experience in juror selection had never been integrated with the list-matching theory developed in other fields. Search soon disclosed an extensive (and highly statistical) literature describing the theory and problems of matching lists of names and for removing duplicates from them.

This report thus attempts to serve a dual purpose. First, it makes specific recommendations to the Superior Court of San Diego County. In addition, it attempts to bring together aspects of the theory and background of list matching to support and illustrate its recommendations. Although the specific recommendations made to San Diego County Superior Court may not fit every other court system and conditions, the general topical form and development given in this report should be widely applicable.

The five topical areas to be addressed are as follows:

- Whether to add additional source lists and which lists to add.

- How to establish optimum criteria for removing duplicates, balancing missed matches (keeping duplicates) against mismatches (rejecting good names), and thus maintaining equal probability of selection for all individuals named on any of the lists.

- How to draw a defensible random sample for each venire in a simple cost effective way.

- How to monitor the results of the merging process for good jury management.

- How to deal with the problem of compensating for the lack of residence jurisdiction information on some lists.

2. RECOMMENDATIONS FOR SAN DIEGO COUNTY

Five specific recommendations are made for San Diego County as follows:

(1) Supplement the voter registration lists as a source of jurors names with the driver lists (DMV) and no other lists.

The purpose of this is to attain a satisfactory cross section of the population of San Diego County; that is, to attain necessary "balance" and "inclusiveness." The purpose and procedures will be explained in detail in section 3.

(2) Establish the following criteria for recognition of duplicates between combined lists:

- Exact last name
- Exact first name
- Middle initial, not full middle name
- Birth month and day, not birth year
- Street number or post office box number, not street name

The purpose of this recommendation is to limit the percentages of not-found duplicates, without increasing the erroneous elimination of good names as if they were duplicates. The weaker the criteria, the more duplicates will be found, but the more good names will erroneously be eliminated. This recommendation will provide an optimum balance between these two types of errors in recognizing duplicates, as explained in section 4.

(3) Use the economically optimum methods for combining lists consistent with the available hand-sorting or computer-sorting facilities for recognition and removal of duplicate names. (Either with hand or limited computer facilities available, sampling methods are greatly superior; however, if ample computer time and ample name storage capacity is available in existing computers, full list combination can be used). [1]

---

[1] Kadane, J. B., and Lehoczky, J. P., "Random Juror Selection from Multiple Lists," Operations Research, Vol. 24, No. 2, March-April 1976.

The purpose of this recommendation is to give the court a choice of combination methods based upon cost and time available in San Diego County. The methods will be explained in detail in section 5.

(4) Reduce the number of retained duplicates and monitor the name-matching process by adding two short key questions to the Qualification Questionnaire form for prospective jurors. The key questions are whether the respondent is on the voter list, on the driver list, or on both lists.

The purpose of the questionnaire follow-up is to match prospective jurors' response against the list from which the name was selected, increasing the precision of removing duplicates. This will be explained in section 6.

(5) Provide a means for selecting residents of appropriate jurisdictions for the four municipal courts and the two superior courts of San Diego County by means of proper "geocoding."

The purpose of this recommendation is to solve complicated jurisdictional problems involved in using names which do not have unique jurisdiction identifiers. Proposed solutions and alternatives are explained in section 7.

Material directly supporting the recommendations is presented in sections 3 through 7 as referenced above. Appended material provides a broader view of the subject matter analyzed. This material is as follows:

(1) Appendix A. Experience with the use of multiple lists in Colorado, Alaska, North Dakota, Idaho, and other jurisdictions.

The purpose of this presentation is to illustrate past experience with the combining of two or three lists and to evaluate where list combination has accomplished useful purposes and where it has not.

(2) Appendix B. The North Dakota list combination method recently developed to ease the clerical burden of combining lists where one list is computerized and one list is not.

The purpose of this material is to provide details of format and instructions to use when smaller jurisdictions wish to combine lists.

(3) Appendix C. General demographic description of the voter and driver lists. After each election, the Census Department makes a detailed study of the characteristics of those who voted state-by-state. Some pertinent aspects of these studies are presented. Not much is known about the composition of the driver lists, for they are controlled, revised, updated, and purged of deceased, non-renewals, moved, and so forth, differently in each state.

The purpose of this information is to encourage those who wish to use these lists for purposes of juror selection to find out as much as they can about the characteristics of the lists in their own jurisdictions before they are used.

(4) Appendix D. Methods and procedures for estimating list overlaps as applied in San Diego County.

The purpose of this section is to provide procedures for studying a local situation to find the probable changes in inclusiveness from adding one or more new lists.

(5) Appendix E. Worksheet for determining duplication errors.

The purpose of this material is to provide a worksheet for estimating duplicate rates.

3. MULTIPLE LISTS TO BE USED

3.1 Summary of Recommendation One

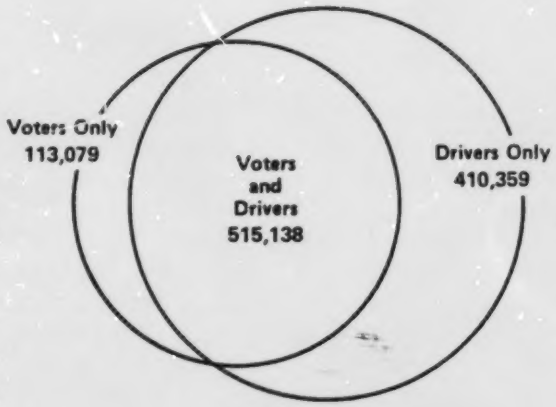
The first recommendation is to select names of prospective jurors from a combination of voter registration lists and driver lists for the following reasons:

- The voter registration list, which contains only 57% of the over 18 population, has been shown to provide an inadequate cross section of vertain demographic groups, especially those living in predominantly Spanish surname tracts.

- The driver list, which contains 83% of the over 18 population, is complementary to voter registration lists and is available. The combined lists should include about 94% of the population, as will be discussed in Section 5 and shown in Figure 1.

- No other lists are considered useful supplements:
  - -- The telephone lists are large but heavily male oriented and involve problems of business listings, unlisted numbers, listed duplicates, and frequent changes. Although the telephone list does appear to add a quantity of unique names not on voter-driver lists, it is suspected that these involve non-persons, internal list duplicates, or missed matches.
  - -- The real property list is relatively small (male names dominate) and is difficult to use because of formats involving multiple ownership and the existence of the great number of financial institution listings.
  - -- The Polk City directory, not countywide, is heavily male oriented and based primarily on residences rather than on persons.

San Diego County, California - 1976	
Registered Voters List	628,217
Drivers Licenses (18+)	925,497
Total	1,553,714
Less Duplicates	515,138
Combined Master List	1,038,576
Population 18 and Over (est)	1,110,783



Voters 11%	Voters and Drivers 50%	Drivers 39%

Combined Master List = 1,038,576 = 100%

Figure 1. Effect of Combining Lists in San Diego County, California

3.2 Discussion

The recommendation to add the driver list to the voter list is based upon the following considerations to be discussed in this section:

- Inadequacy of the voter registration list by itself
- Large size of driver list ensuring inclusiveness
- Complementary nature of driver list providing balance
- Inadequacy of all other lists considered

The lists considered and their characteristics are given in Table 1. The information given on each list is given in Table 2, which is important since any duplicate matching routine requires comparable information. The lack of demographic data, age, sex, race, etc., illustrates the impossibility of determining the effect upon balance of the various list combinations.

Table 1. Characteristics of Possible Juror Source Lists - San Diego County

List	Coverage	Available for Use?	Sex		List Size (1976)	List Cost	Comments
			Male	Female			
Voters	S. Diego County voters	Yes	47.0%	54.0%	628,217	None	Starting 1976, no general purge of list. Female dominated.
DMV	S. Diego County drivers	Yes	54.0%	46.0%	925,497	Cost of tapes	Update every 4 years. Addresses considered 75% reliable. Male dominated.
Welfare	Present and 18 mos. past welfare recipients	No	26.4%	73.6%	47,793 (18+)	$500-$600	Legal impediment. Female dominated.
Telephone	Lists persons in San Diego County	Possible	70.0%*	24.0%**	770,000 listings	Unknown	No economic data. Male dominated.
Polk	Directory for San Diego City	Yes	70.2%*	27.9%*	320,000 (18+)	Unknown	Updated every year by household survey. Not countywide.
Real Property	Owners of property in San Diego County	Yes	50.2%*	49.8%*	260,000	Maybe only tape cost	Names of joint owners difficult to access
Census Projected to 7/21/76	18 and over population	No	51.4%	48.6%	1,110,783	Not available	Shows relative magnitude of other lists.

*Percentage based upon sample rather than published information.
**6% Undetermined

Table 2. Information Coverage on Lists

Information	Voters	DMV	Welfare	Phone	Polk	Real (1) Property
Last Name	x	x	x	x	x	x
First Name	x	x	x	(2)	x	x
Middle Initial	x	x		(3)	(3)	(3)
Mail Address	x	x	x	(3)	x	x
Birth Day	(4)	x	x			
Birth Month	(4)	x	x			
Birth Year	(5)	x	x			
Sex		x	x			
Height		(6)				
Race			x			
Soc. Sec. No.	(7)		x			

Notes:
(1) Many records have multiple names.
(2) First name or initial.
(3) Often not provided.
(4) Mandatory now, but many records do not have.
(5) Mandatory 1976.
(6) Not on list provided by DMV, although DMV has information.
(7) Optional.

3.2.1 Inadequacy of Voter Registration List by Itself

The voter registration list of San Diego County with only 56.6% of the over 18 population is neither inclusive of the population of jury age nor does it have balance among the various census tracts making up the county. Professor Aubrey Wendling has shown that the voter registration is only about 25% of the relative population in Spanish-speaking tracts, whereas it is as high as 91% in affluent, primarily white resident tracts.[2] Data of voter registration levels are tabulated by medium income and ethnic background, as shown in Table 3.

Table 3. Percentage of Registered Voters in Census Tracts for Population 18 Years and Over in San Diego County, 1975

Ethnic Group	Median Income		
	Below $8,000	$8,000 - $11,000	Above $15,000
1/4 to 1/2 Spanish-speaking	27% - 42% (15 tracts)	27% - 49% (4 tracts)	
1/4 to 3/4 Black	30% - 42% (8 tracts)	46% - 64% (4 tracts)	
Mostly White	46% - 58% (5 tracts)	55% - 72% (6 tracts)	69% - 91% (9 tracts)

The inference is that the use of the voter registration lists will tend to over-represent the affluent whites, and thus tend to under-represent the low income blacks and Spanish-speaking groups. This inference might not hold if those registered in the Spanish-speaking tracts were mostly Spanish rather than a broad cross section of both Spanish-speaking and non-Spanish-speaking people. Without further information, however, the under-representation of minority groups does appear to be founded.

---

[2] Differential Voter Registration, A Study Prepared for the San Diego Superior Courts, by Aubrey Wendling, PhD., Department of Sociology, San Diego State University.

3.2.2 Large Size of Drivers List Providing Inclusiveness

In contrast to the voters list of San Diego County which includes 628,217 names (1976) or 56.6% of the population over 18 years old, the drivers list numbers 925,497 of those over 18 which is about 83% of the population.

The drivers license list is so extensive in its coverage that there is a question as to why it might not be used as the sole source list. The apparent reason for not recommending it as a sole list is that about one fifth of those on the voters list are not on the drivers list; it might be considered unwise to deny jury duty to these people who have demonstrated an interest in government by registering to vote. Moreover, these people tend to create a balance that the vehicle drivers list by itself does not have.

In terms of population in age groups between 25 and 64 years old, the drivers list for San Diego County contains 97% of the males and 90% of the females. At ages 65 and over, the percent of females holding drivers licenses drops off to 43%, whereas the percent of men over 65 includes 80% of the population. Below the age of 25, however, this order is reversed with 78% of the females and only 66% of the males in the population holding drivers licenses. This small and relatively lower proportion of young men with licenses is contrary to the intuitive feeling that most youngsters obtain licenses as soon as they can; however, it is a possibility that those in the armed forces and those not registered in their home districts are not counted. Whatever the reason, no quantitative explanations are available. An age and sex breakdown of the drivers license list is shown in Table 4.

Table 4. San Diego Residents Holding Drivers Licenses by Age and Sex*

Sex	Age Group Percentages						
	18-19	20-24	25-34	35-44	45-54	55-64	65+
Male	60.6	67.8	91.0	96.9	104.6	99.7	80.0
Female	61.9	85.9	88.3	90.0	92.9	84.0	42.9

*This information compiled from "Projected Motor Vehicle Registration and Drivers Licenses Outstanding: 1970-1990," Report No. 48, October 1974.

3.2.3 Complementary Nature of Drivers List Providing Balance

The combination of the drivers and voters list can be shown to provide a balance between the age and sex distribution. These two demographic measures are the only ones available on both lists. They show the complementary nature of the lists in known areas but only by inference in those not measured.

In Table 5, the percent of males and females for each of the lists is given. The over 18 census population in 1976 included 51.4% males and 48.6% females for San Diego County in contrast to about 49% male and 51% female nationwide. A sample of 200 names from the San Diego voter registration list verified the 1974 SMSA census figures showing the voter registration to be 47% male. A sample of 570 names from returned Qualification Questionnaires shows that of names taken from voter registration lists, the percent of males decreases further to less than 46%. The percentage of males on the drivers list is larger than the population percentage, and quite opposite from the voters list. The estimated combined list values given in the last column show the results to closely reflect the population as desired.

Table 5. List Sex Distribution (1976)

Source List	Male	Female
Population 18 and over	51.4%	48.6%
Voter Registration	47.0%	52.5%
Returned Qualification Questionnaires	45.6%	54.4%
Drivers List	53.5%	46.5%
Combined Lists (Estimated)	51.3%	48.7%

The age distribution of the five lists is given in Table 6. Compared to the census population estimate, the voter list tends to have fewer young (18-24) with moderate overages in all other age groups, whereas, the driver list has more in the younger group and fewer in the group over 65. The combined list is somewhat closer to the census than either of the other lists, but it does fall short among the 18-24

group and among the over 65 group. Whichever of these three lists is used (voter, driver, or combined), the people 25-64 tend to be over-represented. The qualification process tends to modify this in that more over 65 and fewer in the 25-64 group respond to the Qualification Questionnaire. Qualification may have a similar effect on the combined list, bringing the Qualified Wheel into close correspondence with the census age distribution.

Table 6. List Age Distribution

Age	Population 1976	Registered Voters 1974	Qualification Questionnaire Responses	Driver List	Combined List (Estimated)
18-24	23.6	14.1	13.7	20.0	17.0
25-44	38.0	41.1	31.0	41.7	41.4
45-64	25.4	30.1	34.9	29.1	29.6
Over 65	13.0	14.7	20.4	9.2	12.0
Total	100.0	100.0	100.0	100.0	100.0

3.2.4 Inadequacy of Other Lists Considered

The welfare, telephone, Polk City directory, and real property tax lists were considered in addition to the voter registration and driver license lists. All of these lists are related to the census population in Table 1 for comparison, even though the federal census list itself may not be used for juror selection purposes.

The sex comparison of the lists is the only common demographic characteristic easily available. The driver list is male dominated by about the same degree that the voter list is female dominated. The real property list, although small, is also roughly balanced between the sexes. The welfare list is heavily female dominated in sharp contrast to the Polk directory and the telephone lists which are heavily male dominated. The voter, driver, and telephone lists are relatively large, whereas the welfare, Polk, and real property lists are relatively small in comparison with the indicated population. The welfare is the only list of these which could contribute to the balance of the combined lists.

The welfare list, as of 30 June 1975, contained 52,814 active cases: this included the names of 12,636 male adults and 35,157 female adults, making a total of 47,793 adult names. In addition to sex, information of the list also includes the birthday, social security number, race, and education. Names remain on the list for approximately 18 months after they have finished receiving benefits.

Names on the welfare list were heavily concentrated in the age brackets from 20 through 44. In contrast to 44.3% of the population in this age bracket, 82.6% of the welfare list were of these ages. The problem, however, is that the welfare list cannot be used for the purpose of juror selection due to statutory limitations.

As a list for jury selection purposes, the telephone list has good qualities and bad. The good qualities are that the list is easily available and are published in alphabetical order. However, the list excludes some 25% of individuals who pay for unlisted telephones. It has many duplicates including those who may have two or more addresses, those who have two or more listings (Bob and Robert, John W. and J. Walter, etc.), and those otherwise duplicated. Business names account for 20 to 25% of the total. In a sample of the San Diego directory, the names of individuals were found to be 70% male, 24% female, and 6% unknown, because only initials were given. Despite a search through the Bell System, no characteristics of lists by age, by race, county of origin, or economic status could be found. The telephone list is therefore not recommended.

The Polk City directory covers the city rather than the entire county and contains about 320,000 names of persons over 18. No demographic information is available on the list. The size of the list is smaller than the voter or driver lists, and the demography is expected to be similar. For this reason, the use of the city directory as a further supplemental list is not recommended.

The real property list contains about 260,000 names; however, many are listed under financial institution names representing property in trust. Although the appearance of male and female names in the list is roughly balanced, the first listing is usually male. Use of the on-line terminals to determine the list overlap proved difficult since many property listings could only be accessed by the husband's name. Because of the size of the list and the expected demography, the real property list is not recommended.

4. DUPLICATE RECOGNITION

Any matching routine, whether manual or computerized, requires some criterion upon which to base the decision as to whether two records are duplicates or represent two individuals. Associated with each criterion are two error rates or probabilities which correspond to the two decisions and are:

(1) The probability of rejecting a good name

(2) The probability of keeping a duplicated name

The probability that a list will contain duplicates or that a good name will be rejected in the matching routine is the product of two probabilities, the first being the probability that the matching criterion will fail, that is, reject a good name and keep a duplicated name, and the second being the probability that a pair of names is or isn't a true duplicate. The first probability is a property of the matching criterion and is discussed in this section, while the second probability is a property of the lists considered and related to the percent of duplicates on the list (and the method of merging the lists).

4.1 Decision Errors

The kinds of decisions possible when matching two records fall into one of four categories, as illustrated below in a typical truth diagram:

		Criterion Indicates:	
		Individual Same	Individual Different
True Situation	Individual Same	True Match A	Missed Match C
	Individual Different	Mismatch B	True Nonmatch D

An error is made when either:

(1) The matching criterion does not recognize two records which, in fact, do match (missed match - cell C). The chance of this type of error becomes greater as the criterion becomes more strict and uses more information.

(2) The matching criterion recognizes two different records as representing the same person (mismatched - cell B). Chances of this error occurring increases as the criterion becomes less strict thereby enabling two records to be more easily identified as being similar.

The errors are difficult to measure for they assume a "true situation," which may not be known precisely. To measure these errors a combined list of the voters and drivers was obtained from another California county, which did not have any duplicates removed. The list consisted of 1,740 names. Using all the identifying factors available, which is assumed to be the "true situation," the list was found to have 522 duplicates. A criterion was then established requiring an exact match of last name, first name, middle initial, street number, and street name.

The decisions based on this criterion were:

		Criterion Indicates:		
		Same	Different	Total
True Situation	Same	355	167	522
	Different	5	1,213	1,218
	Total	360	1,380	1,740

The probabilities of making each type of error were then calculated as:

P_{dup} = Probability of a missed matched (duplicate remains)

$$= \frac{167}{355 + 167} = .320$$

P_{rej} = Probability of a mismatch (good name rejected)

$$= \frac{5}{1,213 + 5} = .004$$

A study performed by the Bureau of the Census and the Social Security Administration provided some interesting data for use in determining the dupicate matching error rates.[3] Two lists were available -- the Current Population Survey (CPS), and the Social Security Administration (SSA) Records. Using several criteria, names from the CPS were matched against the SSA. The agreement or disagreement of social security numbers provided verification of the match.

Four match "rules" based on five variables and the allowed tolerances were used:

Variables and Tolerances:

1. Age -- with four-year tolerances
2. Race -- no tolerance
3. Sex -- no tolerance
4. Month of Birth -- no tolerance
5. Surname -- 4 of the first 6 letters of last name must agree

Match Rules:

1. Perfect Agreement -- all 5 variables within tolerance
2. Surname Agreement -- variable 5 only
3. CPS-SER -- 4 of 5 agree within tolerance
4. Potentially Usable -- no restrictions

The data given in the referenced paper was combined to form the probabilities listed below:

		P_{dup}	P_{rej}
1.	Perfect Agreement	.12	.01
2.	Surname Agreement	.05	.02
3.	CPS-SER (4 of 5)	.08	.16
4.	Potentially Usable	0	1.00

[3] "Fiddling Around With Nonmatches and Mismatches," by Fritz Scheuren and H. Lock Oh, Social Security Administration. Paper delivered at the 1975 annual meeting of the American Statistical Association; will appear in the 1975 Proceedings of the Social Statistics Section. Included in Some Preliminary Results From the 1973 CPS-IRS-SSA Exact Match Study, issued 30 September 1975 by the U. S. Department of Commerce, Bureau of the Census.

This shows that as the rule is made less critical so that more duplicates will be rejected, the chance of rejecting a non-duplicated name increases. The best matching criterion depends on the criticality of making either type of error. In terms of their effect on the juror selection process, a missed match error subverts the goal of equal probability of selection because duplicate names are allowed to remain on the list used for selection, and a mismatch error denies a person the opportunity to be selected because the name has been taken off the list. The chance of the missed match error tainting the equal probability of selection, however, should be minimized with the institution of the recommendation using questionnaire verification, a method for identifying duplicate records which have been missed by the checking routine. The optimum matching routine in this application then is one which will assure the fewest number of mismatches. Once a good name has been lost due to this type of error, it cannot be recovered.

4.2 Types of Record Errors

The matching criteria used to determine duplicate records for estimating the San Diego voters and drivers list overlap (see Appendix D) was based upon the information available in each list. The information available is:

	Voters	Drivers
Name:		
Last	x	x
First	x	x
Middle	Initial	x
Address:		
Home number	x	x
Street name	x	x
Birth:		
Day	25% *	x
Month	25%	x
Year	10%	x

*Percent of records which contain information.

The basic matching criteria to manually determine a true match were as follows:

(1) When birth month and date information existed on voters list, the following must agree:

a. Last name
b. First name
c. Middle initial
d. Birth day
e. Birth month

(2) When birth month and date information did not exist, the following must agree:

a. Last name
b. First name
c. Middle initial
d. Home number
e. Street name

All format and minor spelling discrepancies, such as Av., Ave., or Camto Basswood, Cam Basswood, were recorded but ignored for matching purposes. No attempt was made to check for duplicates which may exist within each list itself.

Table 7 shows the areas of information discrepancies between the 164 matched pairs found. An example worksheet to be used for tabulating this information is shown in Appendix E. The three major areas of discrepancies are:

- Availability of records - day and month of birth are missing in approximately 25% of the voting list records; middle initial is missing in 9% of the records.
- Different street addresses - approximately 16% of the matched records have different street addresses but are listed due to same name and birth date information.
- Format and spacing errors - street name format errors (Ave., Av.; Cam, Camto, etc.) occur in 18% of the records matched.

Table 7. Analysis of 164 Pairs of Voter/Driver Records With Matching Last Names

Matching Criteria	Street Name Exactly Same		Street Name Similar But Not Exactly Same		Street Name Different		Total
	Same Birth Day & Month	Birth Date Not Recorded	Same Birth Day & Month	Birth Date Not Recorded	Same Birth Day & Month	Birth Date Not Recorded	
First Name Exactly Same:							
Same Middle Initial:							
Same Street Number	64	34	21	6	0	0	125
Different Street Number	1	0	0	0	20	0	21
Middle Initial Not Recorded:							
Same Street Number	6	2	2	0	0	0	10
Different Street Number	0	0	0	0	5	0	5
First Name Similar But Not Exactly Same:							
Same Middle Initial:							
Same Street Number	2	0	0	0	0	0	2
Different Street Number	0	0	0	0	1	0	1
Middle Initial Not Recorded:							
Same Street Number	0	0	0	0	0	0	0
Different Street Number	0	0	0	0	0	0	0
Totals	73	36	23	6	26	0	164

In this sample only the last name seems to be immune to any discrepancies. This may be due, however, to the fact that this sample contains only last names which start with the letter A, through "Armstrong." A recent study of a combined voter/driver list with most of the duplicates removed by computer revealed a relatively higher duplication rate among list names starting with the letters D and M due to simple spacing (format) errors, e.g., Mc Hugh, McHugh, which caused otherwise identical records to be assumed to be for two individuals. Other studies have also shown last name discrepancies to be a real factor in record matching.[4] These are often minimized by extraction codes or "Soundex" matching systems. However, these are not felt necessary in the San Diego situation.

4.3 Recommended Matching Criteria

Agreement of the following criteria is recommended for identifying duplicate records from the voter and driver lists:

- Last name
- First name
- Middle initial (where record exists)
- Birth month and day (where record exists)
- Street number or post office box number
  (Street name is not included due to format difficulties)

The following format considerations are also necessary:

- No name or number should contain internal blank spaces
- Obvious errors in the records (such as incorrect zip or out of county, town, or city, non-Alpha names or inconsistent numeric sequences should be checked or rejected).

In these criteria the greatest amount of reliable information is used in order to minimize the chance of a mismatch.

---

[4] "Results of Research Into the Methodology of Record Linkage," David M. Nitzberg, in Record Linkage in Medicine.

"A Computer System for Medical Record Linkage," M. R. Hubbard and J. E. Fisher, in Record Linkage in Medicine, ed. by E. D. Acheson, E. & S. Livingstone, Ltd., 1968, p. 165.

In using these criteria, the assumption was made during the sample checking that records match if the last name, first name, middle initial, day, and month of birth are the same, even when street address differences have been dropped. The court should investigate the validity of this assumption by using the information provided in the qualification questionnaire and by sampling the names rejected.

Based on the results of the 164 matched pairs, the expected percent of true matches which will be missed is 17.7% (29 of 164). This includes 1.8% whose first names will differ slightly and 15.9% of the true matches which will be missed due to the record's different street numbers.

The criteria will miss 17.7% of the duplicates. In order to determine the percent of duplicates which will remain in the merged list, it is necessary to introduce the second probability -- that is, the probability of two names being duplicated when the criteria are applied. The criteria are applied every time two names are compared. The number of comparisons is based on the routine used, and is approximately $n_1 + n_2$, where n_1 and n_2 are the number of names on each list. Using the San Diego data as given in Figure 1, if the probability of compared names being duplicated is $\frac{n_d}{n_1 + n_2}$, where n_d is the number of duplicated names, the probability is $\frac{515,138}{1,553,714} = 0.332$.

The probability of a duplicate remaining in the merged San Diego list is therefore $0.177 \times 0.332 = 0.059$. The merged list will be about 6% larger than a duplicate-free list and will contain about 6% duplicates.

The probability of a good name being rejected is almost impossible to measure without actually merging and counting the errors based on manual screening of computer determined duplicates. However, it can be estimated using the data previously given. The following probabilities of accepting a duplicate and rejecting a good name indicate the expected inverse relationship, which their product verifies.

P_{dup}	P_{rej}	$P_{dup} \times P_{rej}$
0.32	0.004	0.0013
0.12	0.01	0.0012
0.05	0.02	0.0010
0.03	0.16	0.0048
0	0	0

The recommended criteria are expected to have a probability of retaining a duplicate of 0.177. Over the range of 0.32 to 0.05, the product is fairly constant and can be used to estimate a P_{rej} of 0.007. The probability that a comparison will be non-duplicated is 1 - 0.332 = 0.668. The probability of a good name being rejected is 0.007 x 0.668 or 0.005.

The duplicate level of 6% will be reduced by the questionnaire checking routine to a much smaller value which will be determined by such factors as juror response errors and listing errors. This shows the unique value of the qualification questionnaire checking routine to reduce the percent of duplicates while retaining the small chance of a good name being rejected.

5. METHODOLOGY FOR THE COMBINING OF LISTS

5.1 Introduction

Once the need for utilizing two or more lists is apparent, the actual method of combining the lists comes into focus. The most direct way of doing this (and the most expensive and tiresome), whether done by computer or manually, is the direct combination of all lists into a single alphabetical or otherwise ordered list from which all duplicates, within reason, have been removed. The intent of such a list is to provide one listing or one card or token for each eligible person such that each one has the same and equal chance of being selected. However, there are problems involved in the removal of duplicates as discussed in Section 8 and in the method for the combining of the lists, which is discussed in this section.

5.2 Direct Combination of Lists

The combining or merging of two lists is a common data processing task. This merging, subject to a matching of the names to remove duplicates, is an added complication. The technique is basically as follows:

(1) The entry on each list is reformatted into a common format upon which the comparison for duplicates can be based. This information is called the key.

(2) Each list is ordered based on the information in the key. If the key begins with the names of the person, the ordering is alphabetical.

(3) The lists are simultaneously scanned, starting at the lowest key value or the top of each list, and a comparison of the entries from each list is made. If the comparison indicates a duplicate, the name is written to the combined list file and the next name is considered. If the names are different, the one of lowest value is written to the combined list file, and the higher value name is kept and compared to the next name on the list from which the lower value name was taken.

The result of this process is a single merged list with the level of duplicates determined by the name match criteria used and the degree of duplication of the lists. This list is then sampled, using some random method to generate a list for use in the qualifying process.[5]

This technique can be used to merge any number of lists, whether the process is a successive merging of two lists until all the lists are combined or whether the combination is done in one step using the technique discussed.

The computer time necessary to perform this method can take several hours. However, it is usually done only once a year at most. If jurisdictions desire a more up to date source list, the process could be performed more often, providing the constitutent lists are updated. Regardless of the method of combining the lists, certain data should be obtained, such as the number of duplicates found and the constituent sizes of the lists. Samples of the duplicates rejected also can be used to monitor the accuracy of the duplicate matching routine.

5.3 Random Selection Without Full List Combination

Techniques for sequentially sampling and checking from several lists to arrive at a random sample with no duplications have been described in the operations research literature, as previously cited. These techniques do not require that the lists be combined, only one list need be scanned for duplicates. If the list most easily scanned is then chosen properly, the task of achieving a random selection is simplified.

The following illustrates the methods of direct combination and random sampling without direct combination. In Figure 2, lists 1 and 2 are given. The combined list of 21 "names" is shown, as are the 9 "names" which are duplicated. The direct combination method described in the previous section would generate such a combined list.

---

[5] "A Proposed Method for Adding Holders of Drivers Licenses as a Source List for Creation of the District of Columbia Court's Master Jury Wheel", Resource Planning Corporation, Washington, D. C. (undated).

List 1	List 2	Combined List	Duplicates
A	A	A	A
B	C	B	F
D	E	C	I
F	F	D	L
G	H	E	M
U	I	F	P
K	J	G	Q
L	L	H	R
M	M	I	T
N	O	J	
P	P	K	
Q	Q	L	
R	R	M	
T	S	N	
U	T	O	
		P	
		Q	
		R	
		S	
		T	
		U	
15	15	21	9

6 Unmatched List 1 Names	9 Matched Names or Duplicates	6 Unmatched List 2 Names
28.5%	43.0%	28.5%

Composition of Combined List = 21 Names = 100%

Figure 2. Example of Combination of Lists

The method for random selection without full combination is illustrated in Table 8. This table represents method #4 of the 5 methods given in the Kadane-Lehoczky paper.

(1) A random sample is taken from each list given in Figure 2 in proportion to its size. Since the lists are of equal length, 15 names, the samples are also equal and a random sample of 5 names is selected from each list and shown in Table 8.

(2) List 1 is considered the "primary list" and all 5 names are defined as "good." The 5 names selected from List 2 are compared to the entire List 1, and the duplicate names (A, L, and Q) are rejected. The remaining names (H and J) are combined with the first five to obtain the random sample of 7 names.

This method achieves the same result as a direct combination followed by a random sample by sampling first and then comparing one sample to the entire other list, and rejecting duplicates. This rejection of duplicates insures that the duplicated names are given only a single chance of being selected, i.e., on the primary list. Either list may be the primary list, with the choice usually based on the list easiest to check or the longer list.

Table 8. Random Sampling From Lists 1 and 2 To Get Random Sample of Combined List

Samples of Five Names		Random Samples From Combined List
List 1	List 2	List 1 Primary
D	A x	D
G	H	G
P	J	H
Q	L x	J
T	Q x	P
		Q
		T
5	2	7

For the sake of further illustration, the voters list and the drivers list are considered. In this second example, the technique is as follows:

(1) Indicate the drivers list as List 1 because it is usually larger than the voters list, is available on computer tape, and is accessible at an on-line terminal.

(2) Designate the voters list as List 2 since it is generally smaller in size and composed of many different section listings, although alphabetical within sections.

(3) In this example, shown in Table 9, the drivers license list contains 900,000 names and the voter registration list contains 600,000 names.

   (a) Select 5%, or 45,000, names at random from the drivers list and accept all of these as valid names (unless there are internal duplicates, which should be removed). Do not check these against the other list.

   (b) Select 5%, or 30,000, names at random from the voters list and check each one of these against the entire drivers list. Approximately 82% duplicates will be found, and these approximately 24,600 names should be dropped. The remaining 5,400 names are added to the 45,000 drivers names to form a combined list of 50,400 names.

(4) The 50,400 names retained as a result of this selection and checking procedure should be randomized since they may retain the order of the lists.

Table 9. List Merging Samples

Item	List 1 (Drivers)	List 2 (Voters)
Number of Names	900,000	600,000
Random Sample	45,000	30,000

If more or less than 50,400 names are desired, then the percent sampled from each list may be increased or decreased as necessary. The important factor is that the percentage remain constant for both lists, i.e., each sample is proportional to the size of the list. Since the overlap of the lists is approximately known (i.e., 82%), the results are predictable within limits. However, this method does not yield an exact number or the same number each time.

If the overlap of the lists is not known, cross checking of a small sample of each list will provide an estimate of sample sizes needed for the purpose.

5.4 Combination of Three or More Lists

In combining three or more lists as a source for jury selection, the principles and procedures are basically the same as when two lists are combined. The first, or background step, is to study the available lists with respect to their overlap. This can be made by sampling as described under the combination of two lists. The lists should then be ordered in terms of size and/or ease of checking. The importance of studying the lists in terms of size and ease of checking for duplicates arises from the checking procedure in which all names selected from List 2 will be checked for duplicates only against the entire List 1; those names selected from List 3 will be checked for duplicates against the entire List 2, and then against the entire List 1, and so on for as many lists as may be used. Such checking is necessary in order to retain equal probability of selection for each name on the combined list, that is, to preserve the randomness of the selection.

Checking for duplication of those names selected from a sample of one list against only the sample from the other list, a shortcut in combination that many jurisdictions have been or might be tempted to adopt, does not produce a random sample from the combined list, and hence should be avoided.

The detailed sampling procedure is now outlined as follows:

(a) Select from List 1 a number of names and consider all of these to be valid.

(b) Select from List 2 a number of names in the same proportion to the size of that list, and check these selected names for duplicates in the entire List 1. Any duplicated names should be rejected and crossed off from the names selected from List 2; if any of these names happen to have been selected from List 1, they should be retained with List 1. The non-duplicates from List 2 are then added to the names selected from List 1 to form the combined list.

(c) Select from List 3 (and from Lists 4, 5, etc.) a number of names in the same proportion to its size. Each name selected from List 3 must be checked against the entire List 2, and then against the entire List 1 to remove duplicates. The names found duplicated on List 2 should be crossed off from the names selected from List 3 and rejected. Those remaining are checked off against List 1, and again duplicates are rejected. The remaining names are added to the names from List 1 and to the non-duplicated names from List 2 to form the combined list. The names from List 1 are never checked against the other lists. The names from List 2 are not checked against List 3 or higher lists, and so forth.

6. ERROR RECOGNITION USING QUESTIONNAIRE RESPONSES

Courts using multiple lists are seldom aware of the level of duplicates remaining on the list or the probability of rejecting non-duplicated names. The ususal response is, "If we send two questionnaires to the same person we tell them to return only one." The first suggestion to a court already using multiple lists would be to examine the "non-duplicated" list for duplicates, and then ask that the computer program count duplicated names and print out a sample of the pairs to determine if non-duplicated names are rejected.

As the criterion for determining a duplicate becomes weaker (in order to reject more suspected duplicates) the probability of rejecting non-duplicates increases. The balance is usually to accept a few duplicates rather than reject good (non-duplicated) names. Courts accept the fact that a few duplicates remain, although the percentage has been found to be as high as 10% in one jurisdiction for certain portions of the list.

The second recommendation suggests that the voter list and driver list be merged by using the method given in section 5. A less satisfactory alternative is to combine both lists into a single list. In either case names should be coded as to whether duplicated, unique to the voter list, or unique to the driver list. In order to further remove duplicates as well as to monitor the duplicate removal procedure, the qualification questionnaires should be coded with this information. In addition, two key questions should be added:

(1) Are you a registered voter in San Diego County?

(2) Do you have a California license to drive a motor vehicle?

Correct answers by the citizen to these questions indicate whether the respondent is listed on both lists, is on the voter list only, or on the driver list only. When compared to the selection code from the source list selection, the answers provide for further duplication removal and error monitoring.

The checking procedure upon receipt of the qualification questionnaires would be as follows:

- Divide the returned questionnaires into two groups, those that are unique to the voter list (which should be about 10% of the names) and all others being unique drivers, and drivers and voters (about 90% of the names).
- Of those questionnaires from unique voters, the responses to questions (1) and (2) should be "Yes", "No", respectively, which is the correct response. Those responding "Yes" to both questions indicate an error in response or in computer matching and should be rejected.
- Of those questionnaires from unique drivers, or voters and drivers, those answering "No" to the second question are rejected. All others are accepted.
- Tabulate those rejected and determine error source:
  - Computer did not find duplicate
  - Computer rejected non-duplicated name
  - Name incorrect or not on driver list
  - Name incorrect or not on voter list
  - Error in response by citizen

While all these error sources cannot be uniquely measured they can be estimated to monitor the performance of the selection system. Some small rate of errors will inevitably occur. If found to be excessive, the errors can be used to determine the proper remedy to improve the system. If the errors are found to be small and acceptable, then a periodic sampling to monitor and verify the error rates is all that is necessary.

The questionnaire checking method is illustrated in Figure 3, and is parallel to the usual selection technique. The source of a person's name is carried with his name on the record and indicated on the questionnaire by either a Code "1" or Code "12", as shown in the figure. When the questionnaires are returned they are screened prior to determining the qualifications of the person. This screening, which can be rapid, should only yield a few questionnaires if the selection process is operating properly. These rejected questionnaires are then arrayed according to the source of the name and the citizen's response. The various combinations are shown in Table 10.

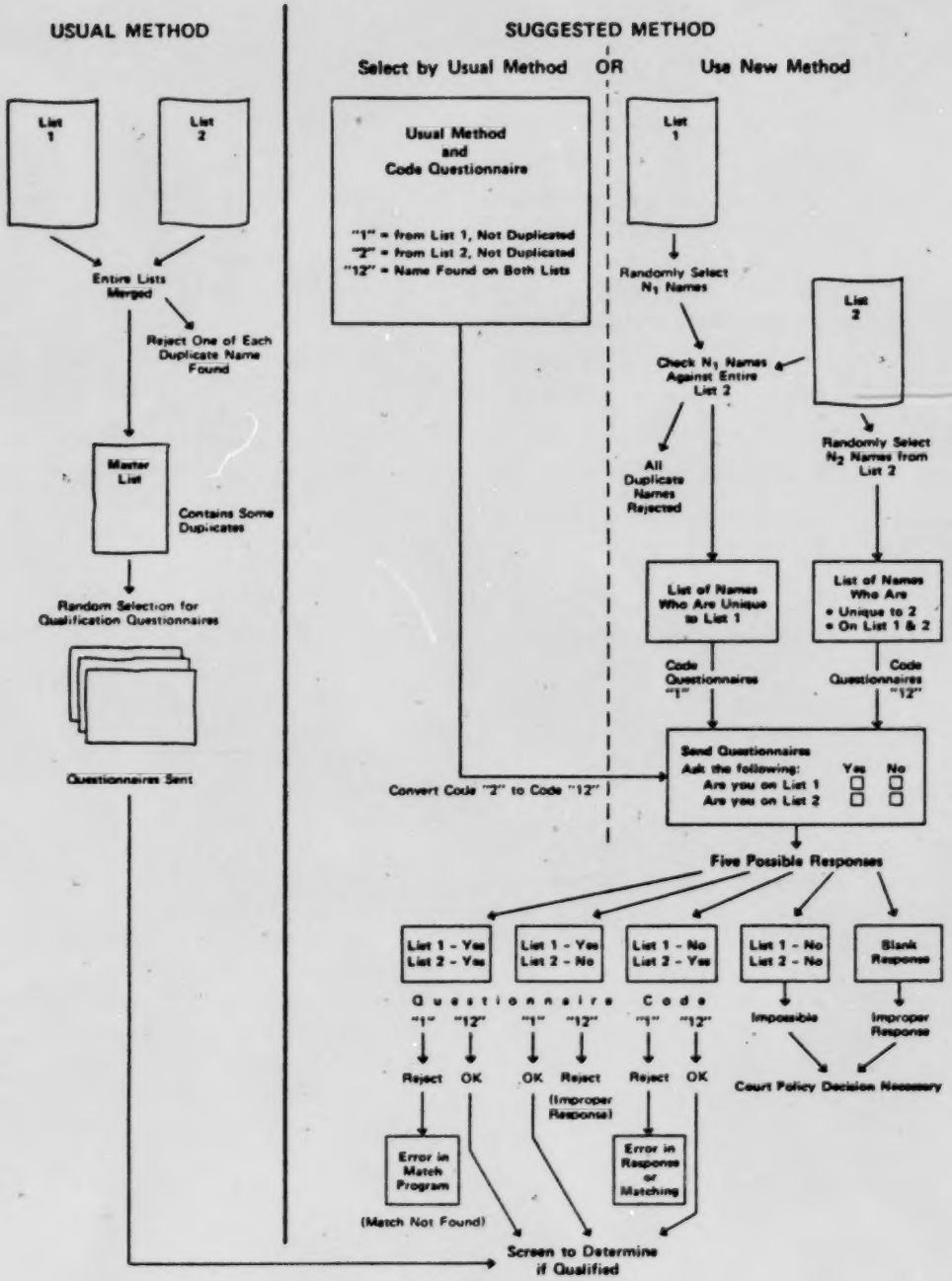

USUAL METHOD
SUGGESTED METHOD
Select by Usual Method
OR
Use New Method
List 1
List 2
Entire Lists Merged
Reject One of Each Duplicate Name Found
Master List
Contains Some Duplicates
Random Selection for Qualification Questionnaires
Questionnaires Sent
Usual Method and Code Questionnaire
"1" = from List 1, Not Duplicated
"2" = from List 2, Not Duplicated
"12" = Name Found on Both Lists
Randomly Select N_1 Names
Check N_1 Names Against Entire List 2
All Duplicate Names Rejected
Randomly Select N_2 Names from List 2
List of Names Who Are Unique to List 1
List of Names Who Are • Unique to 2 • On List 1 & 2
Code Questionnaires "1"
Code Questionnaires "12"
Send Questionnaires
Ask the following:
Are you on List 1
Are you on List 2
Yes
No
Convert Code "2" to Code "12"
Five Possible Responses
List 1 - Yes List 2 - Yes
List 1 - Yes List 2 - No
List 1 - No List 2 - Yes
List 1 - No List 2 - No
Blank Response
Questionnaire Code
"1" "12" "1" "12" "1" "12"
Reject OK OK Reject Reject OK
(Improper Response)
Impossible
Improper Response
Court Policy Decision Necessary
Error in Match Program
(Match Not Found)
Error in Response or Matching
Screen to Determine if Qualified

Figure 3. Qualification Questionnaire Checking Methods

Table 10. Various Combinations of Voter and Driver Lists

Source of Names		Response to Questionnaire: Name from Voters List — Voter not Driver	Name from Voters List — Voter and Driver	Name from Voters List — Driver not Voter	Name from Drivers List — Voter not Driver	Name from Drivers List — Voter and Driver	Name from Drivers List — Driver not Voter
Voter	Voter not Driver	True	B Reject Name	Incorrect Response	NP	NP	NP
	Voter and Driver	A	True	A	NP	NP	NP
Driver	Driver not Voter	Incorrect Response	B Reject Name	NP	Incorrect Response	B Reject Name	True

A = Mismatch, good name rejected
B = Missed match, duplicate not found
NP = Not possible

7. GEOCODING

Selection of names for jury duty in most courts is usually made from a group of residents of the jurisdiction. In cases where non-residents may appear in the group, there is a means of removing names. This editing is usually done by removing the obvious non-residents from the list of people sent qualification questionnaires and then by including a residential question on the questionnaire.

Within San Diego County are two Superior (state) Court locations and four Municipal Court locations. The following shows the jurisdiction from which each court calls jurors:

Superior Court, San Diego	-	entire county
Superior Court, Vista	-	5th Supervisoral District
Municipal Court, San Diego City	-	entire county (shares pool with San Diego Superior)
Municipal Court, South Bay	-	South Bay Judicial District
Municipal Court, El Cajon	-	El Cajon Municipal District
Municipal Court, North County	-	North County Judicial District

San Diego Municipal Court has merged their pool with that of the Superior Court who uses a countywide call. The supervisoral districts are election districts for the County Board of Supervisors. Jurors are called to service in the Vista Superior Court for the 5th Supervisoral District only. The judicial districts were set to define the jurisdiction of the municipal courts. Both supervisoral and judicial districts are defined by a listing of the smaller election precinct divisions. There is no overlap of election precincts nor is any precinct cut by a district. At present, the separation of jurors into their respective jurisdictions is accomplished by a computer precinct to district sorting program of the qualified list.

As the drivers list does not contain precinct or district information, its inclusion as a source list for potential jurors in San Diego County raises the problem of obtaining this information. One possibility explored was to see if the district or precinct could be defined by zip code area. The ideal situation would be that zip code area boundaries correspond exactly to judicial and supervisoral boundaries. A comparison of zip code and district area boundaries shows this not to be the case. Table 11 shows the extent of the overlap in the districts.

Table 11. San Diego County Zip Code Overlaps

Jurisdiction	Number of Zip Code Areas	Zip Code Overlaps	Percentage of Overlap
County	103		
Judicial Districts			
South Bay*	15	4	27%
El Cajon*	27	11	41%
San Diego	55	NI**	29%
North County*	26	8	31%
Supervisorial Districts			
1	26	NI**	
2	36	NI**	
3	29	NI**	
4	20	NI**	
5*	36	17	47%

*Jurors selected from this jurisdiction for a specific court.

**Not Important: Only separate districts which provide jurors solely for one jurisdiction are pertinent.

Data obtained from County of San Diego Registration of Voters District Index File, 7/30/76.

The 5th Supervisoral District has the largest overlap of zip code areas at 47%. This means that of the 36 zip code areas which lie within the boundaries of the 5th Supervisoral District, 17 parts of those areas also lie outside of the 5th district. This similar situation holds true for all judicial and supervisoral districts although it has impact on juror selection in only the 5th supervisoral and South Bay, El Cajon, and North County judicial districts. Many persons who reside in these overlapping zip code areas can be assigned districts if their records match with voting records. Otherwise, their address must be checked manually against an address-precinct list. It is estimated that the maximum number of checks which would have to be made is 1,000 per year.

Another possibility investigated was to use census tract information for locating the precincts of the driver records. Previous experience in San Diego as stated by the county data processing personnel showed this method using the census "ADMATCH" technique to be a problem.[6] Although voters precincts are defined by census tract, there is no other way of relating street addresses to census tracts, and the driver list does not contain census tract information.

The most desirable solutions from an administrative standpoint would be for the court to allow persons who reside outside of a particular judicial or supervisoral district to serve as jurors in the court within that district or to modify current district boundaries in line with the zip code areas.

---

[6]U. S. Bureau of the Census, Census Use Study: Geocoding With ADMATCH, A Los Angeles Area Experience. Report No. 14, Washington, D. C. 20233. See also Reports 1 to 13.

APPENDIX A

BACKGROUND INTO THE USE OF MULTIPLE LISTS

May 1977

1. BACKGROUND INTO THE USE OF MULTIPLE LISTS

At various times in the past, personal discretion has dictated the source of names for juror selection purposes. For instance, in 1880, the California statute indicated that "from those on the assessment roll of the previous year, suitable persons, suitable to serve as jurors -- of fair character, of approved integrity, and of sound judgment" will be selected. This may not mean that the assessment roll was used as the source, for in 1951 this assessment roll language was deleted, leaving only more general requirements of good character. Practice rather than statute suggests that many lists were used to find such names, for in 1954 (People v. White, 43 Cal. 2nd 740, 278 P. 2d 9) the court ruled that a jury list is not condemned because it is complied from directories, such as , "Who's Who, The Blue Book, Social Register, or Club Lists ---."

Use of these many lists was not for the purpose of achieving a wide cross section of the pupulation; rather it was for ease of listing those assumed by the selectors to be "suitable persons -- of sound judgment." The problems in using those multiple lists were not recognized, for there was no implicit cross section or equal probability of selection requirement.[1] The fact that those listed on many lists had many times the probability of selection as those listed on only one, and that those not so listed had no chance of being selected, was not considered germane. However, the Uniform Service and Selection Act of 1968 recognized that random selection from voter registration lists might satisfy the newly developed concept of selection from a wide cross section of the jurisdiction. From that time on, the use of the voter list became more widely used among the states as the source of jurors' names. Even so, use of the assessment rolls still exists in the state of Montana, and many counties in Virginia still use the key man system. The widening use of the voter registration list revealed its limitations in some jurisdictions; not all eligible people register to vote, nor do various segments of the population (minority groups, women, etc.) register to the same extent.

---

[1] Kairys, D., Kadane, J. B., Lehoczky, J. P., "Source Lists for Jury Selection," California Law Review, Vol. 64, No. 4, July 1977.

Although the voter registration list has obvious value as a method of conveying a sense of participation in government and a means of prescreening as to many of the juror qualifications, many recognized its limitations as a cross section from which to draw. The Uniform Service and Selection Act of 1968, in addition to adopting the voter list as the primary source of names, recommended supplementation with other lists. In its Model Code, the National Conference of Metropolitan Courts in 1973 also suggested that voter registration lists be supplemented. As one of the first states to adopt the Act, Colorado began merging lists and has been using multiple lists for years. North Dakota, another state to adopt the Act, manually supplemented their voter list. In addition, other states have supplemented voter lists with others: Alaska added state income and fish and game license lists; Kansas added state census lists; Allegheny County, Pennsylvania, added telephone, welfare, and property lists.

Whereas this increasing use of additional lists may help to over come the weaknesses of the voter registration lists, the resultant characteristics of the merged lists are seldom examined. In order to examine these lists, several concepts are necessary.

"Balance" in a source list, as implied in the Uniform Jury Service and Selection Act, relates to its being "a fair cross section of the population of the area served by the court (Sec. 1)." The Act (Sec. 2) implies that the characteristics considered are, "race, color, religion, sex, national origin, or economic status." Other characteristics such as, age, minority status, and other identifiable classifications may be included in the concept of balance. Balance thus suggests a one-to-one correspondence between the source list and the population with respect to all these classifications that have been considered in the past and which might be considered in the future. The extent to which non-correspondence, with respect to a single characteristic, defeats the concept of balance is not easy to measure.

"Inclusiveness" implies a numerical relationship between the source list and the population considered -- the larger the source list the more inclusive it is. Thus, if a list included everyone in the population, it is totally inclusive. Adding lists together nearly always increases inclusiveness, unless the list added is totally contained in the first. Adding lists may not improve balance, for adding the 'Who's Who' to the 'Blue Book' list, although increasing inclusiveness, may not increase and may even further distort the cross section (i.e., balance).

2. EXPERIENCE IN THE USE OF MULTIPLE LISTS

2.1 Uniform Act

In 1970 the National Conference of Commissioners on Uniform State Laws drafted the Uniform Jury Selection and Service Act. The Act was approved by the American Bar Association on February 7, 1972. Section 5 of the Act provides that:

> 5. [Master List]
>
> (a) The jury commission for each [county] [district] shall compile and maintain a master list consisting of all [voter registration lists] [lists of actual voters] for the [county] [district] supplemented with names from other lists of persons resident therein, such as lists of utility customers, property [and income] taxpayers, motor vehicle registrations, and drivers licenses, which the [Supreme Court] [Attorney General] shall initially designate the other lists within [90] days following the effective date of this Act and exercise the authority to designate from time to time in order to foster the policy and protect the rights secured by this Act (Sections 1 and 2). In compiling the master list the jury commission shall avoid duplication of names.

The Uniform Act has been adopted with very slight variations by Colorado, Idaho, North Dakota, Indiana, and Mississippi. The intent of the Act is that, "A citizen shall not be excluded from jury service on account of race, color, religion, sex, national origin, or economic status;" and in Section 1 that, "All persons selected for jury service be selected at random from a fair cross section of the population of the area served by the court . . ."

The Act intends, therefore, that the "fair cross section" be defined by the six characteristics enumerated, but it does not give the Supreme Court or the Attorney General specific instructions for implementation. The Act does not provide guidance as to how the supplementary list or lists will be selected, nor does it indicate the information necessary to make that decision prudently.

Our review of actual courts suggests that this decision has been made rather blindly -- in fact, sometimes supplementary lists that have been chosen are almost completely duplicated by the principal or voters list. Nowhere has been found a good or complete demographic profile of the lists used, nor of the possible alternative supplementary lists. We do not find a feedback of such information from those actually combining the lists to those making the initial decisions. Information of this kind could be obtained from the component lists or from the combined lists with relatively small samples. In fact, since the prospective jurors drawn at random from the lists are a random sample of the names, a demographic profile including their "race, color, religion, sex, national origin, and economic status" might be quite easily assembled from qualification information to ensure the intent of the Act has been achieved.

Finally, those who make the decisions to supplement the voter list may not recognize the enormous amount of clerical or computer effort involved in an exact or strict interpretation of the Act, for the Act implies that the source lists must be combined into a single alphabetical list from which all duplicates have been removed, and then randomly sampled to provide a Master Jury Wheel. This interpretation in some North Dakota jurisdictions involves the handcopying of names from several lists of nearly 50,000 each, the ordering of the lists, the selection of a Master Jury Wheel of about 1,000 while copying and alphabetizing it along with associated clerical and matching operations in order to select 105 jurors every year. No wonder the clerks ask, "Is this enormous effort worth the result?" (See Appendix B). Similar efforts take place with computer programs. The result is that an enormous amount of paperwork is committed by those who make decisions to combine lists without understanding the consequences. Unfortunately, those who do the work generally do not evaluate the results nor do they make any evaluation information available to the decision maker.

2.2 Experience Under Uniform Act

Colorado, North Dakota, and Idaho adopted the Uniform Act in 1971 with very minor modifications, and have been enriching the voter lists with driver licenses and other lists. In Colorado, the lists are combined by computer as a centralized state operation with the lists furnished to the local jurisdictions. In North Dakota the lists are combined by hand in each of the counties. In Idaho some counties have computerized operations; others combine the lists by hand Last year the Office of the State Court Administrator acted to obtain the driver list and to prepare labels for each of the counties, thus avoiding the previous difficulty of having 44 counties each negotiate with the Department of Law Enforcement. Some Idaho counties also use the list derived from local utility company files.

Although the form is different among the states and among the jurisdictions in North Dakota and in Idaho, the principles are quite the same among them, as follows:

(1) All prepare a single alphabetical list, known as the Master Jury List, from which exact name - exact address duplications have been removed. This list is computer printed or typed.

(2) All then use a random start-fixed interval method to select a substantially large random sample from the Master Jury List, which is then called the Master Jury Wheel. The Master Jury Wheel is small -- usually one or two percent of the Master Jury List. Qualification Questionnaires are usually sent to all those on the Master Jury Wheel.

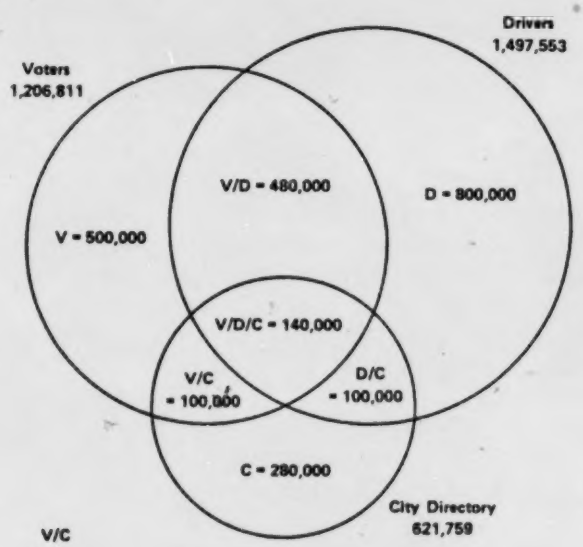
(3) Citizens are summoned at some later date from those found to be qualified, as prospective jurors are needed. The qualified list tends to age as people move, die, or become infirm.

(4) The combined list (that is, the combined Master Jury Wheel) is in all instances found to be larger than the corresponding census estimate of the population 18 years and older. Three to five percent of this difference is made up of those less than 18 on the driver list The excess must be drivers no longer residing in the county, those who have moved, and undiscovered duplicated names.

The situation in Colorado is typical of the difficulty in determining the precise operational effects. The State of Colorado, operating under the Uniform Act since 1971, has been combining three lists, which in 1975 were 1,200,000 voters, 1,500,000 drivers, and 600,000 city directory names. The computer matching of names from the three lists developed a combined list of 2,400,000. This indicates that approximately 900,000 duplicates have been removed from the three lists. A sample carefully examined by hand disclosed 9% duplicates.

The Federal Court for the District of Colorado combined two of these lists, the same 1,500,000 drivers and the 1,200,000 voters starting in 1975, netting a combined list of 2,100,000. However, the criteria for matching names is quite different in the federal computer processing operation than in the state, so that the 2.4 million state total cannot be compared directly to the 2.1 million federal total; each of them is in excess of the over 18 population estimate for 1975 of 1.8 million. Figures 1 and 2 illustrate the combination of the lists in these Colorado courts.

Colorado State Courts - 1975	
Registered Voters List	1,206,811
Drivers Licenses	1,497,553
City Directory	621,759
Total	3,326,123
Less Duplicates	964,860
Combined Master List	2,361,263



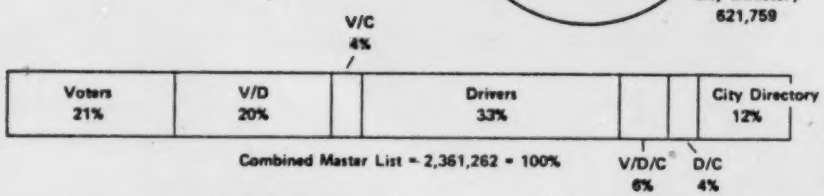
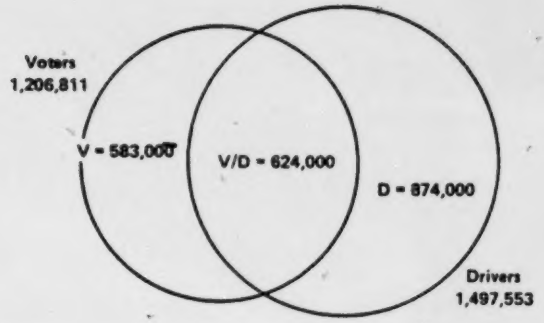


Figure 1. Effect of Combining Lists in Colorado State Courts

United States Court for the District of Colorado - 1975	
Registered Voters List	1,206,811
Drivers Licenses	1,497,553
Total	2,704,364
Less Duplicates	623,604
Combined List	2,080,760



Voters 28%	Voters and Drivers 30%	Drivers 42%

Combined List = 2,080,760 = 100%

Figure 2. Effect of Combining Lists in United States Court for the District of Colorado - 1975

The state list is 300,000 names larger than the federal list. If this difference could be solely attributed to the addition of the city directory list of 600,000 and none to the differences in matching techniques, then using the city directory as a third list might be quite worthwhile in that half of its names are added. This assumes, however, that the names on the state list are as valid as those on the federal. If, however, the addition in names to the state list arises primarily from increase in duplicates, then the addition of the city directory might be practically worthless.

In Ward County, North Dakota, as in most of the other small counties of that state, the driver list is manually combined with the voter list. The voter list contains 21,320 names, and the driver list contains 24,680 names. The lists contribute 25% and 35% unique names, respectively. Both lists are relatively useful in reaching greater inclusiveness.

After the list of 33,000 names is carefully assembled by hand and alphabetized, a jury wheel of 700 is selected at random. A qualification form is addressed to this list of 700. Generally only three panels of 35 jurors each are used during the year. The clerk in this court quite correctly pointed out that combining the lists seemed to involve a great deal of clerical work every two years in view of the small number of jurors selected.

Ada County, Idaho, combines a list of 78,940 registered voters with 90,557 drivers to form a master wheel of 131,630. The computer matching list and first names and initials, as well as the first five characters of the address, eliminates 37,867 duplicates in the matching process. The voter list is observed to have 41,073 unique names whereas the driver list contributes 52,690 unique names. The two lists are quite complementary for the reason that the percentage of duplicates is smaller than that observed in other jurisdictions. After the final file is established, a random sample of 10% is selected to form a Jury Wheel of 13,163.

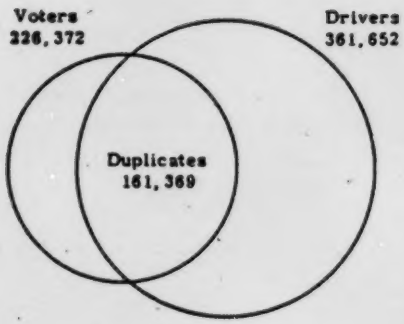
2.3 Other Experience in the Use of Combined Lists

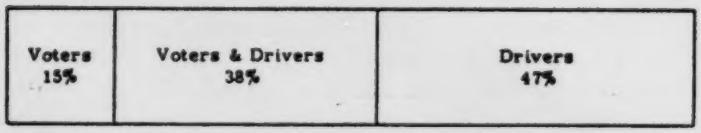
Other jurisdictions, including many in California, Alaska, and Kansas, have joined two or more source lists for juror selection. In San Mateo County, California the court has combined voter registration and driver license lists by computer in much the same way as Colorado. A difference, however, is found in the characteristics of the list, for here the drivers constitute nearly 85% of the combined list in contrast to 72% in Colorado. Moreover, the voter list in San Mateo County is only 53% of the combined list in contrast to 58% in Colorado. Only 15% are unique voters in San Mateo, in contrast to twice that percentage in Colorado. These details are shown in Figure 3. These results raise the question of how much is added by the use of the voter list. If the demographic profile of the driver list is no different from that of the combined list, then in San Mateo and in other similar situations there is virtually no gain from the addition of the voter list; the driver list might be used independently.

San Mateo, California - 1975

Registered Voters List	226,372
Drivers Licenses	361,652
Total	588,024
Less: Duplicates	161,369
Combined List	426,655*
Population 18 and Over (est)	(390,000)

*5.4% duplicates remain, estimated





Combined Master List = 426,655 = 100%

Figure 3. Effect of Combining Lists in San Mateo, California

Alaska combines the voter list. the fish and game (hunter and trapper) lists, and the Income Tax List to form their juror source list. The details of the numbers on each list and the number of duplications among the lists are given in Figure 4. The excellent records kept make possible an analysis of the intersection among the lists.

The number of unique names contributed by a list divided by the number of names on the list provides a measure of efficiency in adding that list. Thus the following is given for the 3 lists:

List			Efficiency
Voters	=	100/168	60%
Fish & Game	=	54/116	48%
Income Tax	=	43/121	33%

This suggests that the Income Tax List is the least efficient because the great portion of its names are duplicated, whereas the voter list is most efficient because fewer names are duplicated.

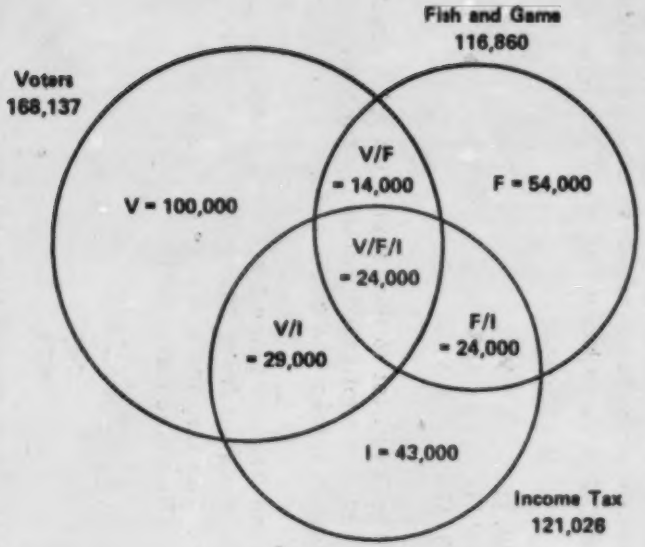
If only the voter and the fish and game lists were used, the combined list total would be reduced from 289,910 by the number of unique names on the Income Tax list and would thus be 246,696. Given these two lists, these new efficiencies are derived:

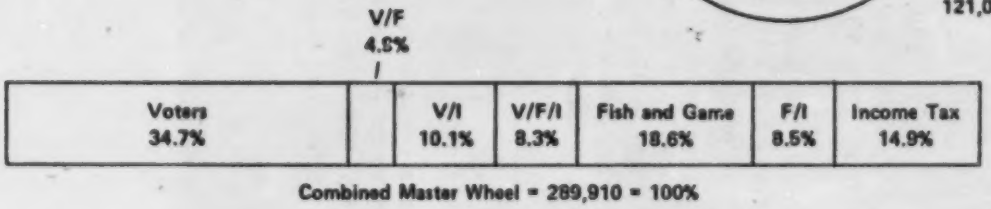
List	Efficiency
Voters	78%
Fish and Game	67%

Thus, both lists contribute to the combined list and complement each other. If only voter and Income Tax lists were used, the combined list would fall to 235,989, and the efficiencies would be 66% and 55%, respectively, indicating that these lists do not complement each other as well.

State of Alaska - 1976

Registered Voters List	168,137
Fish and Game	116,860
Income Tax	121,026
Total	406,023
Less Duplicates	116,113
Combined Master Wheel	289,910





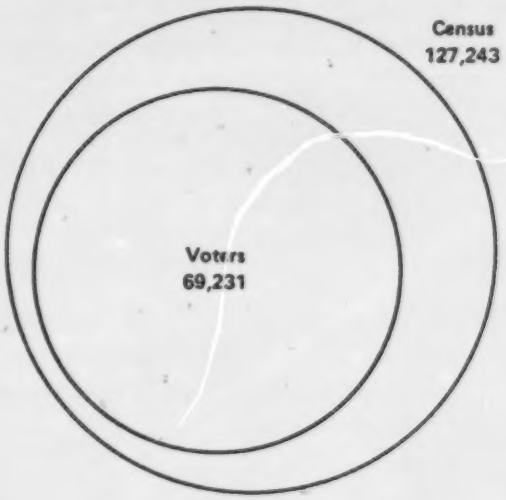
Combined Master Wheel = 289,910 = 100%

Figure 4. Effect of Combining Lists in Alaska

Wyandotte County, Kansas, combines the state census with the voters list as required by statute with the results shown in Figure 5. The combined list is not as large as the census total, indicating that some of the census names were edited out in the duplicate checking program which is an automated merger and a manual check. The result is that the voters list is completely contained in the census list and adds nothing to the combined list. The efficiency of the voters list is 0% and 46% for the census. These lists do not complement each other and should not be used together.

Wyandotte County, Kansas - 1975

Census (State) List	127,243
Voter Registration List	69,231
Total	196,474
Less Duplicates	71,219
Combined List	125,255



Census List Voters	Census List Non-Voters
55.2%	44.8%

Combined List = 125,255 = 100%

Figure 5. Wyandotte County, Kansas, 1975 (Population)

In one New York county, the voter registration list is used for juror selection for the first half of the year and the drivers license list is used for the second half. This practice avoids the complex problem of combining the lists, but it introduces a serious problem in the probability of selection, for those names appearing on the two lists (the exact proportion being unknown) have approximately twice the chance of being selected as those names appearing on only one list. This practice of sampling from one list and then from another does not provide equal probability of selection to persons named on the lists.

A similar alteration in the probability of acceptance arises in those instances in which only a sample is taken from the voters list and another sample from the drivers list, duplicates appearing only in the samples being eliminated. For in this situation, as in the alternation of the lists, a name appearing on both lists over a number of years has about twice the probability of selection as a name appearing on only one list. A method for avoiding this inequity while still taking only samples from both lists is described in section 6.3 of this report.

2.4 Summary of Experience

While the methods used in the courts studied are varied and questionable, the results have provided the desired effect of increasing the size of the source list. Courts using multiple lists have not been challenged and many feel they have prevented an eventual challenge had multiple lists not been used.

What is missing is the jury system management which can verify the results of the merging of the lists, either in demographic terms or in equal probability of selection terms. This report has attempted to bring these results together to provide some guidance to interested courts and to introduce new techniques which can reduce the burden of combining lists existent in many courts.

APPENDIX B

This appendix contains the Bird Associates' memorandum of 29 October 1976 on the use of the Kadane-Lehoczky method for random selection of names from several lists. The memorandum's recommendations were followed in North Dakota, which resulted in a savings of several thousand dollars in Burleigh County. The method also was presented to the clerks of court and jury commissioners at their annual meeting in April 1977.

29 October 1976

MEMORANDUM

1. INTRODUCTION

Under the Uniform Selection and Service Act, Statutes of North Dakota, Chapter 27-09, the Master List for juror selection is defined as the list of actual voters supplemented with other lists. The supplement used is the list of licensed drivers as supplied by the Highway Department. While other states use a wide variety of secondary lists from State Income Tax and State Census, to Hunters, Trappers, and Fishers Licenses, the usual second list, as in North Dakota, is the drivers list. Drivers license lists are relatively bias-free and highly inclusive of the target population. For instance, the driver list in Burleigh County contains about 96% of the population over 18, while the list of actual voters (poll books) contains only 65%. Furthermore, while 47% of the drivers do not vote, only 21% of the voters do not drive. Thus the combined list of 31,000 (about 96% of the population over 18) is composed of about 13,000 drivers who do not vote, 4,000 voters who do not drive and about 14,000 who both vote and drive.

Bird Engineering-Research Associates, under contract to American University, is studying many problems that arise in the use of multiple lists. Some of these problems are discussed in A Guide to Juror System Management (Section 2. 1. 1).

2. PROBLEMS OF COMBINING LISTS

In states which have both lists in computer format, the combining of the list is more easily accomplished, but in North Dakota where the voters list is not in computer format, the key-punching of the entire list of voters would be a huge task. In addition, every 2 years the entire list would have to be reentered rather than just updated, as could be done if voter registration were used.

In Burleigh County, North Dakota, the manual effort to combine the voters and drivers lists takes about a month and requires an additional $3,000 for clerical help. Several clerks in North Dakota have voiced their concern about the magnitude of the effort which is called upon to yield only a few hundred jurors.

3. METHODS OF SELECTION

An article appeared recently in the operations research literature by J. Kadane and J. Lehoczky of the Carnegie Mellon University titled, "Random Juror Selection from Multiple Lists." In that article they develop several methods for randomly selecting names from several lists without actually first combining the lists to form the master list. Their methods go directly to the generation of the master wheel from the source lists. The method maintains an equal probability of selection for all names, fulfilling the intent of the statute, while saving time and money associated with generating the master list. In order to compare the current method with the method to be recommended, both lists are given.

3.1 Current Methods

(1) The list of licensed drivers is obtained from the Highway Department. In Burleigh County this list contains about 27,000 names (1974).

(2) The poll books from the 61 precincts are copied and contain about 18,000 names.

(3) Each name (all 18,000) in the poll books is manually checked on the driver list. If found, the name is "lined out" on the driver list. About 14,000 names are lined out.

(4) The resultant names not "lined out" on the driver list (about 13,000) and all of the names of actual voters from the poll books are sequentially numbered to form the Master List of about 31,000 names.

(5) Using the "key number" method (refered to in Section 2.1.2 of A Guide to Jury System Management as random start-fixed interval) a random selection of approximately 2,000 names from the Master List is made. The 2,000 names are indicated as the Master Wheel and qualification questionnaires are sent to them.

(6) Jurors' names are selected as needed from those found qualified.

---

[1] Operations Research, Vol. 24, No. 2, page 207, March-April 1976.

3.2 Recommended Method

(1) The number of actual voters in the precinct poll books is determined. For illustration purposes, the number voting in 1974 was 18,000.

(2) The number of licensed drivers in the county is obtained from the Highway Department. In 1974 there were 27,000 drivers in Burleigh County.

(3) Using the "key number" method, a random sample of 1,200 names is obtained from the poll books. No duplicate checking is done in this operation. This number 1,200, is 6.7% of the voter list.

(4) Request from the Highway Department the following:

- A random selection of 1,800 names, those names by definition are "good." This is 6.7% of the driver list.
- Check using the computer on-line terminals the 1,200 names selected from the poll books. If the voter is a licensed driver, his name is "lined-out." If the voter does not have a license, his name is good. This process will yield about 21% "good" names or about 250 names out of the 1,200 checked.

(5) The "good" names combine to total about 2,050. These names are designated as the Master Wheel.

(6) Qualification Questionnaires are sent to the names on the master wheel and jurors are selected from those jurors qualified, as before.

4. ADVANTAGES OF THE RECOMMENDED METHOD

(1) No manual duplicated checking is done by the courts saving a great deal of clerical time.

(2) The duplicate checking done by the Highway Department is computer assisted and will take an estimated 10 seconds per name. For the 1,200 names, this is about 3 1/3 hours.

(3) The method reported used by several courts of selecting a number of names from each list and eliminating only the duplicates found in the samples does not provide equal probability of selection and must not be used. This doubles the probability of selection for the persons who are named in both lists.

5. EXPLANATION OF RECOMMENDED METHOD

The recommended method gives equal weight to the lists since the number of names selected is based upon the ratio of the size of the two lists, that is:

	Number of Names on List	Names Selected	Percent of Names Selected
Drivers	27,000	1,800	6.7%
Voters	18,000	1,200	6.7%
Ratio	3:2	3:2	

This method does not yield an exact predictable number of names, for the number derived from the second list (as in Step 4b) may be slightly more or less than the 250 indicated. If sufficient names are obtained or more are needed due to increased jury activity, then an additional sample may be taken with the only requirement being that the ratio of names taken from each list be the same. The percentage selected from the driver list should be the same as from the voter list.

If the check for duplicates yields greater or less than the estimated 21% from the voter list, the number of good names is greater or less, but the equal probability of selection is not affected.

The following explanation will further illuminate the recommended method. Each name on a list is either unique or duplicated on the other list. That is, names on the voter list are either:

A -- Voters Only

B -- Duplicated on the drivers list

The names on the driver list are either:

C -- Drivers Only

D -- Duplicated on the voter list

The names in group B are the same as in group D. In a random selection from the combined list, these names should be allowed to enter the selection only once -- either as B or D. The method described includes names from groups C and D (step 4a) and from group A, but rejects all names in B (step 4b).

The current and recommended methods can be compared another way using the A-D nomenclature described.

The voter list contains 18,000 names of which 4,000 are unique and 14,000 are duplicated, as shown below:

Voter List

A Voters Only 4,000	B Duplicated on Drivers 14,000

The driver list contains 27,000 names of which 14,000 are duplicated, as shown below:

Driver List

C Drivers Only 13,000	D Duplicated on Voters 14,000

Using the current method, either group B or D is eliminated by checking each name on one list against the other list. The result is a single list, called the Master List, which would contain:

A Voters Only 4,000	B or D Duplicated Drivers or Voters 14,000	C Drivers Only 13,000

A random sample based on the estimated number of names desired is then taken from A+B+C; that is, from the 31,000 names. (The statute gives a minimum value). The resultant names are called the Master Wheel.

The recommended procedure combines the steps of merging lists with duplicates removed, and sampling the resultant list. The sample from the voter list (poll books) included those names in groups A and B. The computer checking against the Highway Department records removes those in B. The sample of the driver list is a sample of C and D. Thus, this recommended procedure also samples from A+B+C. Since the number of names from each list is in the ratio of the size of the individual lists, each person has an equal chance of being called just as in the current method. This fulfills the Legislative Intent of the statute without generating the Master List.

6. IMPLEMENTATION SCHEDULE

The recommended method could be used for the generation of the Master Wheel following the November 2, 1976 election in Burleigh County. Rather than using the computer to determine the duplicates, the list already prepared by the Highway Department could be used. This would still save clerical time in that only about 1,240 name matches (and not 18,000) will be required.

The recommended method is completely compatible with the current metho; therefore, use need not be statewide, but used only by those courts who understand the method and recognize the savings possible. Once the method has been tested in Burleigh County, the procedures could be extended to other interested courts with computer assistance for the name matching.

APPENDIX C

DEMOGRAPHIC CHARACTERISTICS OF VOTER REGISTRATION LISTS AND MOTOR VEHICLE DRIVERS

May 1977

DEMOGRAPHIC CHARACTERISTICS OF VOTER REGISTRATION LISTS

In the 1974 Congressional elections, 62.2% of the eligible population was registered to vote; some 47.7% actually voted. The Census Department, after this election as after previous elections, made a survey of the demographic characteristics of the 62.2% who were registered to vote and thus, by inference of those not registered.[1]

Since the voter registration lists are widely used as a single or partial source of jurors names for jury selection, the demographic characteristics reported by the census indicate the cross-section characteristics of the eligible population reached by these lists. High and low percentages of registration were reported to be among the following groups, as measured against the average 62.2% registration nationwide:

High Voter Registration		Low Voter Registration	
Persons 45 to 64 years	- 73%	Persons 18 to 34 years	- 47%
Residence in North Central United States	- 67%	Residence in South or West	- 60%
White	- 64%	Spanish Origin - 35%; Negro	- 55%
College Graduates	- 76%	4 Years or Less of Education	- 52%
Annual Income $25,000 up	- 78%	Annual Income $5,000 or less	- 52%
Professional and Technical Workers	- 78%	Laborers	- 48%
Men	- 63%	Women	- 62%
		Moved Within Past Year	

When the Census inquired into the reasons why people failed to register, "not interested" was by far the most frequent reason given. Those demographic groups lowest in voter registration were generally highest in the "not interested" reason given, except for those of Spanish origin where "not a citizen" was the most frequent response.

---

[1] U.S. Bureau of Census, Current Population Reports, Voting and Registration in the Election of November 1974, Series P-20, No. 293, (Washington, D. C.: United States Government Printing Office, April 1976).

Of some other ten reasons cited for failure to register, the fear of being called to jury duty as a result of registering was not included, possibly being hidden under "other" or "did not know." The groups generally higher in voter registration tended to give more plausible reasons for not registering, such as "moved recently" and "physical disability." The Census estimated that 17% of the people moved within the last year, and that only 37.5% of these registered to vote.

These general characteristics of voter registration lists both reflect and help to explain the great variations among states, among cities, and among counties within the states. Among the states, Virginia has the smallest percent registered (54.0%), whereas Indiana has the highest (69.6%). The Miami, Florida Standard Metropolitan Statistical Area (SMSA) with its large new Cuban population has the lowest percentage (49.9%) registered among metropolitan areas reported, followed closely by New York SMSA (51.1%), where volatility helps explain the low percentage. Voter registration in states and metropolitan areas reflects the general population composition: stably situated white affluent white-collar citizens more frequently register than others.

County registration percentages reflect similar characteristic differences within the states. In Maryland, for instance, where average registration is 64.1%, the most affluent county, Montgomery, reaches 79%, where a less affluent Prince Georges is 52%, and farming St. Mary's dips to 51%.

Voter registration lists also expand as national elections approach. For instance, in San Diego County the voter registration list was 675,000 in July 1976, increased to 700,000 in September, and is expected to reach 840,000 by the time of the elections in November 1976. The voter registration in Prince Georges County, Maryland, reached a high of 240,000 after the 1972 elections, fell to 215,000 in 1974, and is climbing upward to 265,000 as the 1976 election approaches.

Jury rolls derived from voter registration lists thus have important limitations with respect to both criteria of inclusiveness and of balance. Voter registration lists do not cover all of the eligible population, and the part they do cover varies widely from one population entity to another. Voter registration covers just about a quarter of the eligibles in Spanish speaking tracts of San Diego, but nearly all of the eligibles in parts of Montgomery County, Maryland. These

differences in coverage of voter registration lists contribute to their lack of balance, for it is apparently the Spanish speaking, the Negro, the less affluent, the blue collar laborers, the young, and the perpetually moving people that refrain from registering. Balance, or lack of balance, is not uniform among the population entities from which jurors might be drawn, for it can be demonstrated that those jurisdictions with lower voter registration reach a less representative cross section of the population. Presumably, the quarter of the population registered in the San Diego census tracts are the English speaking, more affluent, more stable, more upwardly mobile citizens of that community. The small percentage not reached in the Montgomery County tracts are primarily the young, the less affluent, the poorly educated, and the laboring men of that community. Lack of balance, although not absent at any level, decreases as inclusiveness increases.

The adequacy of voter registration lists for jury selection purposes can be answered only with respect to the situation within each court jurisdiction. In some jurisdictions, the voter registration lists may afford sufficient balance through high participation of a relatively homogeneous population; in others, the balance may be so inadequate as to exclude important cognizable groups.

Thus, the question of adequacy of the voter list must be faced in every jurisdiction. The safest protection for any court is to know where it stands with respect to the population cross section from which it draws jurors. The court should collect relatively small samples of demographic information from the source list from which names are selected for prospective jurors. Since small samples are now selected at some point for juror qualification purposes, a summary of the information generally requested on the questionnaire is all that is needed. If this information is collected both on the list queried and on the jurors actually reporting for service, a court can reach important decisions on the basis of routine summaries. The decisions will include those with respect to the adequacy of the source lists, with respect to statutory and ruled exemptions and exclusions, and with respect to the excuse policy followed. This valuable information can be obtained, assembled, and used with the only cost being covered by the deletion of much unnecessary paperwork now clogging most jury selection systems.

DEMOGRAPHIC CHARACTERISTICS OF LICENSED MOTOR VEHICLE DRIVERS

In 1975 the Federal Highway Administration estimated that 129,814,873 persons were licensed to drive motor vehicles by the fifty states and the District of Columbia. This total is 81.4% of the driving age population, a percentage that has been steadily increasing one or two percentage points per year since the data were recorded. 54.4% of the drivers were male, the percentage of males being more than 50% in all the states. A predominance of male drivers is indicated in all age groups, but the percentage of males increases in age groups 55 and over. In the larger states, approximately 94% of the males of driving age population and 70% of the females are licensed to drive.

Since drivers license lists are now used in three states and in many counties as a partial source of jurors' names, a description of the demographic composition of drivers license lists is given in general terms as follows:

High Percentage of Drivers	Lesser Percentages of Drivers
Rural states - Ga. - 98.0%	Urban states - N.Y. - 64.1% D.C. - 59.9%
Males all ages - 90 to 100%	Females over 55 - 30 to 40% Low income city youth

The effect of wage levels, per capita income, topography and climate of the state, race, color, religion, national origin, and other demographic factors is not anywhere reported. One might surmise on the basis of general observation, that the percentage of registered drivers is lower among Spanish speaking and low income women of all ages but especially those over 55, whereas for most men of all ages, the drivers license is a necessary adjunct of employment both for transportation and for identification.

All states include the drivers name, address, signature, and birthdate. All but a few states show sex, weight, and height; some still show color of eyes and hair, many now include a color photograph of the individual as identification upgraded each time the license is renewed. Race is an identifying item in some 29 states. In some states, therefore, some demographic data on drivers may be derived by sampling the lists, but none of this is available on a national basis.

The Federal Highway Administration collects information from the states each year to publish a bulletin entitled, "Drivers Licenses - 19_ _." Only 43 of the states provide a breakdown of drivers by sex and age and the others are estimated from contiguous states. However, the accuracy and currency of the state lists are not known, nor is the amount of duplication within the lists or between the lists from different states known. Anomalies appear to be contained in the state reports For instance, in 1974 Georgia reported a number of drivers licenses equivalent to 98.0% of the population. For males in Georgia in 1974, the drivers licenses exceeded the relevant population numbers by 20%. In 1975, however, Georgia showed a decrease of 414,000 licensed drivers bringing the estimated percent of driving age population holding licenses from 98% to 85%. No explanation of this reported change in number of drivers was available. Similarily, Idaho decreased from 96% of driving age population reported in 1974 to 88% in 1975, a change that may have coincided with the computerization of the list. Alaska, Kentucky, Florida, Ohio, and Wyoming all increased from 1974 to 1975 in the percentage of drivers, although except for the first state with its immigration of young people there is no apparent reason for the shift in driver to population ratio. Wyoming, in fact, reported that the number of drivers licenses exceeded driving age population by more than 1%.

In 1975, all states and the District of Columbia reported to the Federal Highway Administration the total number of driver licenses, but there was no assurance that this number represented the current number of drivers in each state. The Bureau suspected that some states might have had three or four years of expired licenses still carried on their roles. Duplication in drivers licenses among the states may also exist, but this total is not known. Most states try to pick up a prior license in another state before granting a new license, but there is no strict enforcement. Individuals may have two or more licenses, not only in different states, but in the same state under slightly different names (Bob and Robert, etc.), or at different addresses. Some within-state duplication may be avoided in the eleven states that use Social Security numbers as drivers license numbers and in the additional 22 states that include the Social Security number on the license. Nineteen states still do not use Social Security numbers and thus do not have this positive means for checking duplication.

The Federal Highway Administration feels that it has no way of enforcing accuracy in the state records; all it can do is accept what is reported. Seven states in 1975, including New York, were unable to provide a summary of driver licenses by age groups and sex. Some of these states apparently do not have the driver licenses computerized and others complained that they have no funds with which to make the necessary tabulation since, they say, "no one else asks for this information."

The National Driver Register, established by Congress in 1960 under Public Law 86-660, is intended as a voluntary state system for recording drunken and otherwise irresponsible drivers (usually involved in fatal accidents). Other states can check applicants against this list. The register does not provide any means for checking duplication of non-problem drivers among the states, nor does it provide any systematic demographic data on these problem drivers.

As a source of names for jury selection purposes, state driver lists have good and bad qualities. The good qualities are that they cover a large proportion of the population of jury age, thus providing inclusiveness, that they are for the most part computerized (45 to 50 states), that each list entry is one and only one individual, and that they have positive individual identification through Social Security in 33 states, which allows for elimination of within-list duplication.

The poor features of the driver lists for jury selection arise from the possibility that the records may not be reliable because of duplicates within and between states, because of non-removal of expired licenses, and because of lack of control of address changes. These poor features may not be disadvantageous to the lists in support of their primary purpose of ensuring that only qualified people drive. Driver lists have the further disadvantage that almost no demographic information is available as to the cross section of the population which they include. From the present records, aside from small section comparisons as in the San Diego list comparisons reported elsewhere in this report, no information can be obtained concerning the cross section coverage of driver lists by cognizable subgroups.

Any jurisdiction might, therefore, wish to have such cross section information before depending on driver lists as a source of names. This information can be obtained for any jurisdiction by questionnaire sampling of a relatively small sample (500 random names) drawn from the list.

APPENDIX D

DETERMINATION OF LIST OVERLAPS AND EFFICIENCIES

May 1977

DETERMINATION OF LIST OVERLAPS AND EFFICIENCIES

Even after source lists for juror selection are combined, few jurisdictions know the contribution of each list to the total. Courts considering the use of multiple lists should have some estimate in order to determine if the combination is useful. The concept of list efficiency introduced in Appendix A of this report is one measure; the presentation given in Figure 1 for San Diego is another. In this appendix, the method used to determine these values for the San Diego lists will be described.

To determine the overlap between the voter and driver lists, a sample of 200 names was drawn from the voter list and manually checked against the entire available driver list.[1] Results are given in Table D-1, showing that only 18% of the voters in San Diego County are non-drivers. The overlap between the two lists is estimated to be 515,138 records. The efficiency of the lists can be estimated from these figures at 18% for the voter list and 44% for the driver list.

The reliability of these results was then tested by reversing the process and checking a sample of 100 names from the driver list against the entire available voter list. Results are shown in Table D-2 and are summarized along with the voter sample results in Table D-3. The results of the two samples show less than 1% difference in the percentage of overlap measured. This difference appears to be insignificant, and the results of the larger voter sample have been accepted as valid.

Several samples from several portions of the alphabet would have been useful in verifying the value of the overlap. However, the variation expected from such additional samples would not be expected to affect the recommendations or the techniques given in this report.

---

[1] Samples checked were from a complete listing on only "A" through "Armstrong" from both lists. Included in this portion are 18,343 voter records and 28,628 driver records. Matching criteria used is described on page 20.

Table D-1. Voter/Driver Overlap Estimate From Voter List Sample

Category	Sample Size	Percent of Sample	Voters List	Drivers List
Total	200	100%	628,217	925,497
Voter/Driver Overlap	164	82%	515,138	515,138
Voters Only	36	18%	113,079	--
Drivers Only	--	--	--	410,359

Table D-2. Driver/Voter Overlap Estimate From Driver List Sample

Category	Sample Size	Percent of Sample	Drivers List	Voters List
Total	100	100%	925,497	628,217
Driver/Voter Overlap	55	55%	509,023	509,023
Drivers Only	45	45%	416,474	--
Voters Only	--	--	--	119,194

Table D-3. Summary of Sampling Results

Category	Voter List Sample		Driver List Sample	
	Number	Percent	Number	Percent
Total Combined List Minus Duplicates	1,038,576	100.0%	1,044,691	100.0%
Voter/Driver Overlap	515,138	49.6%	509,023	48.7%
Voters Only	113,079	10.9%	119,194	11.4%
Drivers Only	410,359	39.5%	416,474	39.9%

APPENDIX E

WORKSHEET FOR ANALYSIS OF DISCREPANCIES IN PAIRS OF VOTER/DRIVER RECORDS WITH MATCHING LAST NAMES

May 1977

WORKSHEET Date_____

For Analysis of Discrepancies in Pairs of Voter/Driver Records With Matching Last Names

Matching Criteria				Street Name Exactly Same		Street Name Similar But Not Exactly Same		Street Name Different		Total
				Same Birth Day & Month	Birth Date Not Recorded	Same Birth Day & Month	Birth Date Not Recorded	Same Birth Day & Month	Birth Date Not Recorded	
LAST NAME EXACTLY SAME	FIRST NAME EXACTLY SAME	Same Middle Initial	Same Street Number							
			Different Street Number							
		Middle Initial Not Recorded	Same Street Number							
			Different Street Number							
	FIRST NAME SIMILAR BUT NOT EXACTLY SAME	Same Middle Initial	Same Street Number							
			Different Street Number							
		Middle Initial Not Recorded	Same Street Number							
			Different Street Number							
Totals										

RELATED READING

Beauregard, Raymond A., and Kuo, Dr. Shan S., University of New Hampshire. "Evaluation of Identification Factors." Prepared for U. S. Department of Commerce, Bureau of Public Roads, Office of Highway Safety, Driver Performance Division. Washington, D. C.: December 1965.

Kadane, J. B., and Lehoczky, J. P. "Random Juror Selection from Multiple Lists." Operations Research (Vol. 24, No. 2, March-April 1976).

Kadane, J. B.; Kairys, D.; Lehoczky, J. P. "Source Lists for Jury Selection." University of California Berkley Law Review. Berkley, California: September 1977.

Lock Oh, H., and Scheuren, Ftitz. "Fiddling Around With Nonmatches and Mismatches." U. S. Department of Commerce, Bureau of the Census. Washington, D. C.: Social Security Administration, 1975.

Martin, M. P. The Petit Jury System in Alaska. Office of the Administrative Office. Alaska Court System. Anchorage, Alaska: September 1977.

Proceedings of the International Symposium. Record Linkage in Medicine. Oxford: July 1967.

Resource Planning Corporation. "A Proposed Method for Adding Holders of Drivers Licenses as a Source List for Creation of the District of Columbia District Court's Master Jury Wheel." Washington, D. C.: 1975.

U. S. Bureau of the Census. "Voting and Registration in the Election of November 1975." Current Population Reports (P-20, No. 293). U. S. Government Printing Office. Washington, D. C.: 1976.

U. S Department of Transportation. Federal Highway Administration. "Drivers Licenses - 1975." Washington, D. C.: July 1976.

The following documents are available from U. S. Department of Commerce, Bureau of the Census, Washington, D. C.

Reports From the Census Use Study:

General Description. An overview of the development and operations of the New Haven Census Use Study.

Computer Mapping. A report on the mapping of census and local data using several computer mapping techniques.

The DIME Geocoding System. A report on the development of the DIME geographic base file including a description of the file and the edit system, uses of the file, and methods for creating a DIME file.

Geocoding with ADMATCH -- A Los Angeles Experience. A report describing the use of ADMATCH with a variety of local geographic base files such as ACG, DIME, and street tract indexes.

Census Use Study: Computer Program Packages:

DOS ADMATCH: An Address Matching System. A computer program package designed for use in assigning geographic codes to local records using a DIME or similar geographic base file. Includes a users manual and computer programs (written in IBM S/360 assembly language).

OS ADMATCH: An Address Matching System. A new version of the DOS programs specifically redesigned for more rapid processing of large files on medium-size and large computers. Includes a users manual and computer programs (written in IBM S/360 assembly language).

GRIDS: A Computer Mapping System. A computer program package for use on both small- and large-scale computers. It produces within a grid pattern structure density, shading, and value maps. Includes users manual and computer programs (written in Basic Fortran IV).

DIME: A Geographic Base File System. A computer program package for creating a DIME geographic base file. Includes clerical instruction, a computer manual, and programs (written in Basic Fortran IV).

☆ U.S. Government Printing Office: 1978—260-992/2137